Symposium on Big Data and Privacy in Education

TRE

Theory and Research in Education
2018, Vol. 16(3) 280–307

DOI: 10.1177/1477878518815340
journals.sagepub.com/home/tre
SAGE

Learner Privacy in MOOCs and Virtual Education

Elana Zeide
University of California, Los Angeles, USA

Helen Nissenbaum
Cornell Tech, USA

Abstract

The current debate about student privacy issues raised by education technology focuses on how schools information with private vendors. It neglects a related but different trend, namely, the rise of online learning platforms offering learning experiences and credentials directly to users. These Massive Open Online Courses (MOOCs) and Virtual Education providers seek to 'disrupt' the traditional schooling system and position themselves as the next evolution in education. With privacy as the lens, this article highlights problematic dimensions of virtual learning platforms, which fashion themselves as education providers while shaking off the normative and regulatory constraints of traditional educational institutions. Structuring our evaluation around the theory of contextual integrity, we argue that by adopting commercial marketplace norms, these providers undermine core functions and values of education, which include promoting democracy, equal access to opportunity, and self-actualization as well as economic growth. Traditionally, the physical, normative, and regulatory constraints on school information practices created relatively hermetic learning environments. In contrast, Virtual Learning Environments automate instruction, maximize data collection, and codify learning outcomes according to the limited parameters of data-defined metrics and credentials. Because they collect information directly from learners without school mediation, independent Virtual Learning Environment providers fall outside the scope of student privacy regulation and can share information broadly without learner consent or consideration of educational purpose. Our concern is that the new practices risk chilling expression; encouraging narrow viewpoints and filtering out intellectual exploration; exacerbate existing inequities by raising stakes and retaining longitudinal records; and reduce learning to a purely instrumental exercise focused on economic outputs and quantifiable outcomes. MOOCs and Virtual Education providers must go beyond compliance with data collection and use regulation to preserve the values supported by student privacy norms.

Corresponding author:
Elana Zeide, University of California, Los Angeles, CA, USA 90095, USA.
Email: ejz201@nyu.edu

Keywords
Adaptive learning, Coursera, Ed tech, education, edX, e-learning, embedded assessment, FERPA, intellectual privacy, learner privacy, learning environments, Massive Open Online Courses, personalized learning, personalised learning [sic], school, student privacy, Udacity, Virtual Learning Environment, Virtual School

Introduction

Educators and entrepreneurs have long heralded new technologies as transformative, despite repeated cycles of high hopes followed by sobering realities (Watters, 2014, 2015). Once again, reformers' aspirations are ambitious, this time their sights set on harnessing big data, cloud computing, data mining, and analytics to create automatically 'personalized' learning platforms (Carey, 2015). Our article focuses on a cluster of applications within this sphere, commonly called MOOCs (Massive Open Online Courses) and, increasingly, Virtual Learning Environments (VLEs). VLEs provide many functions of traditional educational institutions. They rely on data to deliver instruction, facilitate learner discussion, administer assessment tools, confer credentials, and connect participants with potential employers. VLEs providers seek not only to supplement but also to supplant traditional schools (Byrnes, 2015). Their increasingly important role in the education system makes it crucial to consider the broader impact of information flow on the goals, ends, and values of education itself (MacCarthy, 2014; Zeide, 2017b). Accordingly, we adopt the theory of privacy as contextual integrity (CI) as a framework for understanding and evaluating VLE information practices. We warn that the importation of commercial privacy norms into digital learning systems is inimical to America's pluralistic educational enterprise.

Our article is structured as follows. In the first part, 'Background', describes the rise of independent virtual learning providers and platforms. The second part, 'Student privacy in the educational context: The traditional landscape', presents the traditional information practices and privacy norms in learning environments as reflected in regulation and public rhetoric. Under the commercial regulatory regime, these do not have the presumptions of confidentiality, user consent and access, or educational use that characterize student data norms. The third part, 'Learner privacy in the VLE context: The digital landscape', describes the new information flows and resulting privacy norms of VLEs providing education directly to users. We then situate the information flow in online learning environments' relation to the normative and regulatory constraints of traditional education institutions.

In the fourth part, 'VLE information norms undermine prominent education paradigms', we examine the new information norms, given prominent paradigms of education and their accompanying goals, purposes, and values. Often, student data debates reflect more fundamental disputes about education pedagogy, policy, and purposes (Selwyn, 2014: 125). VLEs promise to provide broad access to education, streamline skill and credential acquisition, align learning content with employer needs, and improve pedagogical efficiency and efficacy. However, education overly focused on economic

goals undermines other important values like democracy, equality, and self-actualization. In contrast to traditional classroom spaces, MOOC/VLE practices include ubiquitous data collection, embedded assessment, competency-based credentials, and extra-educational purposes. Learners lack the practical obscurity of physical settings, traditional consent, access, and amendment rights and the protection of education purpose limitations. The resulting information norms include: (1) Streamlined Academic and Career Attainment; (2) Data-Defined Metrics and Outcomes; (3) Digital Mediation and Data Maximization; (4) Automated and Embedded Assessment; and (5) Private Actors and Priorities. These promote economic efficiency, labor-market harmonization, and career success, but undermine the intellectual privacy, diversity, and experimentation essential to democratic and self-actualization paradigms. They may undercut long-term prosperity and meritocracy as well by discouraging or discarding the outliers whose creativity often creates tomorrow's profitable innovations.

In the fifth part, 'Education versus commercial marketplace: Straddling irreconcilable contexts', we suggest that if VLEs seek to promote similar functions, values, ends, and purposes of traditional education, they must take similar measures to preserve the values supported by student privacy norms in the traditional education setting. If they take on the role of educators, they must also take on the responsibilities to consider learner vulnerability and the negative effects of pervasive surveillance, constant assessment, and indefinite data retention. If VLE providers aspire to conquer the educational context, this uneasy misalignment must be resolved. This involves more than simply regulating user data collection and its use. It requires more fundamental scrutiny of the role VLEs play – and the data they use to teach, assess, and create credentials – not only in learners' lives, but in the emerging private, non-accredited education system.

Background

MOOCs and VLEs

When large-scale online learning platforms initially came to public prominence, they created a mania. The hype started in 2011, when Stanford professors Sebastian Thrun and Peter Norvig offered a free online version of their artificial intelligence class that garnered over 160,000 participants. In the wake of this success, independent platforms delivered open access, free, and on-demand courses online, often partnering with elite higher education universities which provided the course content (Voss, 2013). They became known as MOOCs, an acronym for 'Massive Open Online Courses'. They were 'massive' because they could be deployed at scale with virtually unlimited enrollment; 'online' because they delivered instruction, facilitated communication, and conducted assessment via the Internet. They were 'Open' to all users, originally, at no cost, and they offered learning experiences modeled around the courses taught at traditional higher education institutions. Often, schools used these new platforms to provide online instruction (Young, 2013). A university, for example, would contract an MOOC **provider** like Coursera to deliver recorded lectures, relevant references, and host discussion fora (Kolowich, 2013a). Visionaries and pundits heralded MOOCs as the

future of education, offering lower cost access to quality instructional content at scale (Koller, 2013). *The New York Times* declared 2012 as the 'Year of the MOOC' (Pappano, 2012).

MOOCs were the vanguard of a burgeoning industry of online education platforms hosted by private, predominantly for-profit providers (Ghilay, 2017). Online and mobile learning is now an industry of its own, annually generating billions of dollars (Wan, 2016). The early MOOC providers all 'pivoted' in 2014–2015, altering their business models in response to the financial necessity of raising revenue (Kalman, 2014). Even edX, the nonprofit platform featuring content from prestigious schools like Harvard and MIT, could not remain 'open' without charging fees for certificates (Shah, 2016b).

The 'Big Three' MOOC providers – Coursera, edX, and Udacity – continue to offer some content for free, but charge users for certification and validation completion and performance. They have since thrived. In 2016, MOOCs served 58 million users internationally, providing access to 6850 courses. Coursera is the largest MOOC, providing over 1700 active courses to 23 million registered users. In 2016, it had over 100,000 paid active learners each month. edX serves 1300 courses to 10 million students and Udacity has 4 million learners (*The Economist*, 2017).

The online education market generates over a 100 billion dollars each year and is expected to grow to US$446.85 billion by 2020 (*CB Insights*, 2016). It includes not only traditional 'MOOCs', but a varied array of 'MOOC-ish' pedagogical, technological, and business models. Instead of being Massive and Open, these platforms may limit enrollment to a smaller subset of students. Rather than entire 'courses', many offer modular options oriented around a narrow skillset. Learners' educational experience online may be supplemented by in person instruction or meetings. Some students earn credits or degrees from traditional accredited higher education institutions online (Shah, 2016a).

Many individuals, however, choose more streamlined and less expensive credentials offered independent of the formal education system (Pesare et al., 2015). They earn skill-specific badges, certificates, and micro-degrees in VLEs that range from short coding camp intensives to multiyear programs (Murray, 2017). Employers increasingly accept online learning certificates as valuable credentials (Craig, 2016).

Digital learning platforms now present themselves as substitutes, rather than supplements, to traditional education (Young, 2015b). The edX platform, for example, claims its platform is 'The Future of Online Education' (edX, 2018a). Sebastian Thrun is explicit about his intention for Udacity to replace traditional education institutions. He proposed Udacity as a viable platform and a better alternative to community colleges, brashly forecasting that in 50 years' time only 10 higher education institutions will exist and that Udacity is likely to be one of them (Leckart, 2012). Even if this brash claim does not come to pass, experts predict that acquiring credentials from VLEs will be a crucial component of career success as automation requires workers' 'reskill' (Banks and Meinert, 2016).

Methodology: Contextual Integrity

We apply the theory of CI to examine the normative propriety of MOOC information flow. CI introduced three key concepts into the privacy vocabulary:

> **Contexts:** These refer to social contexts, not formally constructed but, as characterized in social theory and philosophy, natural constituents of social life, routinely embodied in the structures of commercial life and legal domains. Drawn from intuitively recognizable contexts, such as family and home life, politics, healthcare, and education, CI assumes contexts to be defined by several key elements, including paradigmatic activities, roles (or capacities), practices, norms of appropriate behavior, and contextual ends, purposes, and values.
>
> **Contextual informational (privacy) norms:** Among contextual norms, privacy norms govern information flows. As theorized by CI, well-formed privacy norms will specify five parameters: senders, recipients, and information subjects, information types (topics, attributes), and transmission principles. The parameters of actors and attributes range over contextual ontologies, which are distinctive for respective social contexts, if not unique. In an educational context, for example, senders, recipients, and subjects range over people acting in the capacities of teacher, professor, student, principal, department administrator, TA, guidance counsellor, etc. and topics may range over test performance, grades, attendance, classroom behavior, etc. For purposes of this paper, the data subject is almost always the student. Transmission principles condition the flow of information from party to party, for example, *with permission of data subject*, or *in confidence*, as in 'the professor shared the student's grade with his parents with the student's permission'. Informational norms, like other norms, shape people's expectations; norms violated may be met with surprise, annoyance, indignation, protest, or simply helpless resignation.
>
> **Contextual ends, purposes, and values:** Informational norms may affect respective parties ('stakeholders') in various ways, most simply, promoting or impeding their interests, that is, benefiting or harming them. Information flows may also impinge on societal values, such as equality, justice, and various liberties. They may also affect the achievement of contextual ends, purposes, and values–either promoting or confounding them. For example, informational norms enabling (and enforcing) a secret ballot protects autonomous voting for political representatives and, as such, promotes ends and values of democracy. Thus, norms may be evaluated based on their service to respective interests, ethical and political values, and respective ends, purposes, and values associated with respective social contexts. Informational, or privacy norms are morally defensible if they meet the evaluative requirement of interests, ethics, and contextual ends and values.

Contextual integrity is respected when legitimate informational norms are respected; violated otherwise. Because legitimacy of norms depends on how effectively they promote contextual values and purposes, they respond to cultural difference and changes over time. Norms naturally adapt to their surroundings. This does not make a theory of privacy fickle; it keeps it relevant, particularly in the face of rapid developments in digital technologies.

Applying CI to the questions we pose here about VLEs and privacy, we must map personal information flows expected within VLEs and compare these with entrenched information flows in traditional educational settings, the latter guided by explicit law,

local conventions, professional principles, and so on that have evolved. Mapping practices and norms involves mapping flows according to the five parameters. For this article, we will focus on the parameters of actors (senders, subjects, recipients) and information types (attributes).

Having identified discrepancies, a CI analysis involves evaluating competing practices according to purposes and values. Thus, we compare winners and losers in each setting; compare which setting better serves justice or autonomy, and so on; and, finally, compare how well each serves the functions, that is, ends and purposes, of the educational context, as well as values central to it, such as equality, meritocracy, and independence.

Student privacy in the educational context: The traditional landscape

We begin with a brief account of privacy's place within the landscape of education, traditionally conceived (Nissenbaum, 2009). During the past 50 years, the principles concerning appropriate flow of student information have remained remarkably consistent. Long-standing federal law, more recent state regulation, industry self-regulation, and public discourse reflect a sense of school spaces and student records as requiring heighted privacy protection (MacCarthy, 2014; Zeide, 2016a, 2017a). They highlight consistent norms of confidentiality, parental/student rights to access and challenge the accuracy of the content of education records, and education purpose limitations.

Below, we detail the data practices and regulation of student information in paradigmatic physical classrooms, and uncover the student privacy default expectations and norms. Traditionally, human educators and physical schools (1) kept student information confidential absent student/parent consent or school oversight and approval of recipients' legitimate education interests; (2) ensured parents and eligible students could access and contest information in education records; and (3) only disclosed and used student information to further legitimate education interests. Most information about students remained within schools and was used for educational purposes (Cuban, 2013; Zeide, 2016a, 2017a). Federal regulation of education data applies to information about students enrolled in schools that receive public funding, whether directly or through student tuition payments supported by federal aid (Family Educational Rights and Privacy Act (FERPA), 1974; Protection of Pupil Rights Amendment (PPRA), 1978). A newer spate of state student data regulation focuses on primary and secondary schools and the vendors providing them with educational technology (Zeide, 2017a).

Informational norms governing flow and recordkeeping in traditional classrooms

We begin by analyzing aspects of information flow in school environments.

Subjects. We limit our consideration to publicly funded, recognized, or accredited schools which fall under the federal student privacy regime. We also set aside information that these institutions maintained about educators and administrators. Here, we focus on the

paradigmatic setting and data subject: students in physical classrooms or on school premises.

Attributes

Classroom observation. In physical classrooms, most information about students' performance and behavior is not recorded in education records. In the previous era of one-room schools, teachers did not use age or academic grades. Teachers used 'data', to assess student progress and adjust their instruction. Educators cannot capture much information about students' day-to-day interactions and participation in physical classrooms (Zeide, 2016a). Most of this information remained within institutional confines.

Academic performance. Teachers often recorded information about students' performance on quizzes and tests in gradebooks or other personal records. At the semesters' end, they would assess students' performance and submit a grade for the schools' official education records. The limited portability of paper records meant that educators could not share student information widely and without effort (Zeide, 2016a). Student transcripts and degrees included only basic information about student performance and attainment, typically enrollment, grades, instructors, honors, and degrees (Hutt, 2016: 18).

Demographic and behavior information. Schools also traditionally collect high-level demographic, disciplinary, and non-cognitive information about students. Behavioral and disciplinary information, however, was rarely incorporated into students' formal credentials absent academic-related disciplinary infractions (Hutt, 2016: 19–20).

Administrative information. Administrative data include basic contact, enrollment, and identification information and information used to provide non-academic services like busing, cafeterias, and security. Schools can disclose most basic contact information freely under FERPA's 'directory information' exception (34 C.F.R. § 99.31(a)(11) (2015)).

Actors. Whether private or public, most post-secondary schools in America receive federal financial support through direct funding or indirectly via student loan programs. Enrolled students enjoy the heightened protection of FERPA's student privacy regime that, among other things, limits school sharing and outsider access to their personally identifiable student information. It applies to information that VLEs collect about students, whether manually or automatically. Unless an exception applies, schools cannot disclose covered information without consent, either of the parent or eligible students (enrolled in post-secondary institutions or over the age of 18 years). Information recipients cannot not redisclose the information unless it is thoroughly de-identified ('De-identified data', n.d.).

School educators and employees. In traditional educational settings, educators were the primary recorders and recipients of observational information about students. They would share information between themselves to provide better services to students or deal with problematic behavior. Any disclosure about students' day-to-day behavior

and activity to friends and family outside the institution would be relatively narrow and localized (Fitzgerald, 2014).

Contractors and vendors. Schools share administrative information about students as necessary when outsourcing function to outside vendors who provided services like busing and cafeteria management. They do so under FERPA's 'school official' exception, which obviates the need for consent when sharing student information serves 'legitimate education interests' (34 C.F.R. § 99.31(a)(1) (2015)).

Admission boards and employers. In addition to records used for internal administrative or instructional purposes, schools create credentials designed to represent students' academic attainment to outsiders like admission boards and employers (Hutt, 2016: 18). They share these upon the request or with the permission of students or parents (see Zeide, 2016a).

Student privacy norms in traditional education

As the analysis above indicates, schools and policymakers traditionally treat personally identifiable student information with special care. These norms reflect the assumption that the relationship between students and educators is more than commercial: that schools govern data collection and use as part of a special relationship between learners and teachers (Fidanza, 2015; Perez-Kudzma, 2007). This fosters the trust necessary for students to make mistakes, mature, and focus on the task at hand rather than future reputation (Zeide, 2017a).

This stance aligns with education's crucial role in American society and its historical role of serving *in loco parentis*. While educators do not have legally fiduciary duties to their students comparable to lawyers and doctors, they have both legal and tacit constraints on their information practices as reflected in regulation and rhetoric (White, 2007). Courts have long recognized that, given the important purpose of public education and the expansive freedoms of speech and thought associated with the university environment, universities occupy a special niche in our constitutional tradition (see, e.g. *Keyishian v. Board of Regents of Univ. of State of NY*, 1967: 603; *Shelton v. Tucker*, 1960: 487; *Sweezy v. New Hampshire*, 1957: 250; *Wieman v. Updegraff*, 1952: 195 (Frankfurter, concurring)). As noted by Neil Richards (2008), intellectual privacy is a core component of a functioning democracy. The traditional student privacy norms we outline below promote psychologically safe learning environments.

Default confidentiality absent student/parent consent or school oversight. As one researcher notes, '[c]lassrooms are traditionally regarded as relatively private spaces where students can safely explore many areas without exposing their experiences to public scrutiny' (Marshall, 2014). Parents and schools traditionally consider student privacy in terms of disclosure beyond institutional boundaries and educational use, with confidentiality as the norm absent parent/student consent or school oversight (Lu, 2013; Madden et al., 2012).

Parent/student access and ability to contest personally identifiable information. Student privacy norms also emphasize the importance of transparency regarding information practices and providing parents to ensure trust in teachers and schools. FERPA requires schools to allow students or their parents the rights to access and challenge information in students' education records (34 CFR § 99.10 -99.12; §99.20-99.22). Norms permitting parents and student access to and the ability to challenge the content of student records also foster trust. FERPA reassures parents and students that the information in school records is based on accurate, fair, and relevant data. Newer state laws also seek to create governance structures within and outside of schools to ensure adequate oversight regarding school privacy practices (Briones, 2018). California, for example, requires that contracts between schools and vendors include specific data-related provisions and that schools to post contracts with data-reliant vendors on their websites (Student Online Personal Information Protection Act (SOPIPA, 2014)).

Student data only disclosed and used to serve legitimate education interests. Student privacy norms also hold that personally identifiable student information disclosed should only be used for educational purposes. FERPA permits schools to share covered student information absent parental consent when doing so serves 'legitimate education interests' (34 C.F.R. § 99.31(a)(1) (2015)). Newer state laws contain similar provisions that condition disclosure or vendor data use to practices promoting educational purposes (Zeide, 2017a). California law is confined to operators who knowingly provide services used primarily for K-12 'school purposes' to only use information, including creating profiles, to serve such purposes. It defines 'K-12 school purposes' as those

> that customarily take place at the direction of the K-12 school, teacher, or school district or aid in the administration of school activities, including, but not limited to, instruction in the classroom or at home, administrative activities, and collaboration between students, school personnel, or parents or are for the use and benefit of the school. (SOPIPA)

Student privacy norms also emphasize constraint on commercial access to and use of student information. The Protection of Pupil Rights Amendment (PPRA) requires federally funded schools to obtain written consent before administering surveys that include sensitive information like political beliefs or religious practices as well as collection, disclosure, or use of personal information for marketing or sales (20 U.S. § 1232h).

Industry self-regulation adopts similar principles regarding student information (CoSN, 2014). Signatories to Student Privacy Pledge (2016) organized by Future of Privacy Forum and the Software & Information Industry Association (SIIA), for example, testify that they will not sell users' personal information or 'build a personal profile of a student other than for supporting authorized educational/school purposes or as authorized by the parent/student'.

Learner privacy in the VLE context: The digital landscape

The analysis above demonstrated the degree to which student information garners special protection in order to prevent unauthorized disclosure, assure students/parents

of their accuracy, and prohibit disclosure or use that does not further legitimate education interests. We now turn from traditional education settings to the information practices and norms of online and VLEs. Ideally, to ascertain information flows in VLEs, one would have needed direct access to actual information practices. We have settled for a distant second best, studying the published privacy policies of the dominant three VLE providers. Coursera and Udacity have very similar privacy policies – several headings, sections, and phrases are identical. These policies state in very general terms that they record information to provide, administer, evaluate, and improve their services. Even cross-referencing these documents with news articles, interviews, and research articles, it is difficult to determine which practices platforms actually engage in or simply reserve as a future option. This uncertainty is in and of itself a far cry from the transparency and access consistent with student privacy norms.

Below, we detail the data practices and regulation of learner information in MOOCs and other VLEs, to uncover their new privacy norms. Unlike traditional schools, digital education providers do not offer pupils the special protection afforded student records. They do treat learner data as special – for its value as an intellectual and commercial asset.

Informational norms governing flow and recordkeeping in VLEs

We again begin by analyzing aspects of information flow in digital and VLEs.

Subjects. With VLEs providing educational services directly to users, the data subjects are not, legally speaking, 'students'. However, they are still individuals engaged in educational experiences, with instruction, assessment, and credentialing similar to traditional counterparts. The term, 'learners' differentiates them from students, for reasons of idiosyncratic legal definition, but also differentiates them from consumers shopping online or users engaging with social media.

Actors. MOOC/VLE Providers: These are the first parties whose servers collect, store, and process information about data subjects. Learner information is not merely recorded but may be shared with third parties, for example, those who provide additional security and analytics services and those who may find other value in this information.

Contractors. A significant amount of student information passes through third parties that run facility operations, supply data management systems, provide e-textbooks, and increasingly, provide online, adaptive, e-textbooks and courses. VLEs reserve the right to share information with subcontractors, business partners, or affiliates in providing learners with instruction and assessment. VLE providers can also share information with marketers and product developers to discover market needs and adjust their offerings (Prinsloo and Slade, 2015). While VLEs may perform their own analytics to create adaptive environments, most rely on outside learning analytics specialists like Civitas and Knewton. Knewton's founder has boasted:

> We literally know everything about what you know and how you learn best, everything . . . Because we have five orders more magnitude of data about you than Google has. We literally have more data about our students than any company has about anything, and it's not even close. ('Knewton–Education datapalooza', 2012).

Researchers. Some VLEs share information with internal or external researchers who seek to optimize platforms and pedagogy through experimentation (edX, 2018b). Coursera conducts thousands of A/B experiments on users to evaluate pedagogy, design, and metrics that influence student outcomes, using machine learning (Leber, 2012). Researchers are eager to examine VLE data to learn more about 'cognition, metacognition, motivation, affect, language, [and] social discourse' (Herold, 2014). HarvardX and MITx, for example, aim to publicly disclose as much data as permitted under the law to encourage research (Ho et al., 2014). Several VLEs have been working to create an interoperability framework to facilitate cross-platform analyses (O'Reilly and Veeramachaneni, 2014).

Purchasers. VLEs can also use learner data like any other corporate asset. Coursera has contemplated selling student information to employers or advertisers (*The Chronicle of Higher Education*, 2012: 40–41). Coursera's privacy policy indicates it will use nonpersonal information for unspecified 'other business purposes' (Coursera, 2018). It is unclear whether this might include creating targeted advertising or marketing based on learner profiles. edX directly monetized anonymized student data, exchanging it with publisher Elsevier in exchange for free textbooks (Kolowich, 2012).

Attributes. The range of information types VLEs collect, generate, and record is broad and varied, including standard demographic information, formal and informal student output, and performance data. The information collected through these tools is not only broader and more granular than previously available information but contains ancillary metadata (Cope and Kalantzis, 2016: 3–4).

Entered information. VLEs of course collect information users voluntarily and explicitly submit to the platform, such as assignments, answers to tests, and chat room posts (Hollands and Tirthali, 2014). However, the scope of data VLEs collect passively while providing their services is not clear. edX defines 'personal data', but its definition appears to apply only to information learners consciously share with the platform like names, date of birth, and driver's licenses or other government issued identification (edX, 2018b). Coursera and Udacity provides no definition for personal information but instead contrasts it with 'non-personal information', defined as 'information that cannot be used to identify you'.(Coursera, 2018; Udacity, 2018b)

Administrative and learning data. VLEs collect data and metadaeta to authenticate users' identity and track both individual attendance, progress and completion (Cousera, 2018; edX, 2018b). Their policies indicate they do so to provide, evaluate, and improve their services. Through surveys, VLEs may gather demographic information, familiarity with course content, how students learned about the current VLE, and prior online learning experiences (Hollands and Tirthali, 2014).

External data. Besides user interfaces, VLEs may also mine data from other external sources. This might include information about student's learning style, competencies, demographic, geolocation, and social media activity. For example, Udacity may collect and use information from offline or third parties like social media sites (Udacity, 2018b). Coursera's policies also indicate that '[f]rom time to time', it 'may also use additional typical methods of collecting data'. It provides no further information about what those may be, nor whether they are 'typical' according to classroom or commercial privacy norms (Coursera, 2018).

Biometric information. VLEs increasingly collect information about learners through sensor-based technologies and the Internet of Things (IoT) to ensure the veracity of their certifications through remote proctoring applications (Hollands and Tirthali, 2014). Applications like Pearson Vie and Proctor U collect digital, audio, visual, and biometric information to verify students' identities and detect behavior patterns that suggest cheating. This might include biometric signatures, head placement, sounds, eye movement, and facial recognition (Singer, 2015). Coursera, for example, requires learners who want a certificate to provide photo identification that includes names and birthdates and take a typing drill to profile a learner's unique typing patterns. They validate users' identities before and during assessment by comparing this data with a new webcam photo and comparing the stored and real-time typing profiles (Bolkan, 2015; Singer, 2015).

Learner privacy practices in VLEs

Data that VLEs collect from learners do not enjoy the same heightened protection as student data collected within traditional educational institutions. As noted previously, VLE information practices fall outside the purview of FERPA's protection (Young, 2015a; Zeide, 2016a), which only applies to students enrolled in schools that receive federal funds (§ 1232 g(a)(3); 34 C.F.R. § 99.1(a)(1)–(2)). State student privacy laws also focus on information collected and created in traditional school settings (Briones, 2018; Vance, 2016; Zeide, 2017a).

Instead, VLE information practices currently fall under the broader privacy regime governing general commercial transactions and entities, constrained only by their stated Terms of Service and Privacy policies (Young, 2015a). With the notable exception of Khan Academy, most VLEs explicitly eschew the privacy protections traditionally found in school settings. To drive the point home, they assiduously avoid the term 'student' in their policies and on their websites, instead preferring the passive voice and the neutral term, 'users' (Cousera, 2018; edX, 2018b; Udacity, 2018b). edX's privacy policy says 'please note that your education records are protected by the Family Educational Rights and Privacy Act ('FERPA') to the extent FERPA applies'(edX, 2018b). Whatever the intent of this provision, it does not provide learners with the equivalent student privacy protection. It merely states that FERPA protects FERPA-covered information – that is, the information collected from students enrolled at publicly funded educational institutions. Doing so sheds no light on the contours or content of such protection – nor, importantly, does it expand FERPA protection to include learners who use edX platforms independent of schools.

Given this vagary, it is no wonder that few scholars have examined the information norms governing digital learning platforms. In the following section, we expose norms gleaned from MOOC privacy policies, terms of service, and public commentary. Instead of the privacy protection afforded student data, virtual school policies support (1) ubiquitous collection of user information without meaningful mechanisms for learner consent, access, and ability to contest records; (2) constant observation and testing that eradicates academic freedom and intellectual privacy; and (3) maximal disclosure and use for research and revenue-generating purposes.

Ubiquitous collection with minimal user consent, access, and ability to amend. MOOCs say they seek to provide poor, underserved, and nontraditional students access to high-quality instruction, not only in the United States, but also all over the world (New, 2013). VLEs' mission statements echo the ideal of promoting broad access to education as the means to ensure equality of opportunity. edX boasts that its platforms provide education 'for anyone, anywhere, anytime'.

Udacity also focuses on expanding access to education, trying to reach 'people outside the current context of college' (Anders, 2013). Founder Andrew Ng said, 'most people will never have access to a Princeton, Stanford, Cal Tech class . . . But now . . . you can just sign up for one, and it's free' (*PBS Newshour*, 2013). Now, clickstream collection and the possibility of data mining creates a system of 'anytime, anywhere' access, which comes with anytime, anywhere surveillance. Digital media enable an unprecedented level of monitoring and data capture.

Digital education platforms and programs constantly collect and analyze users' actions and answers to assess learner progress in real time. In education, as elsewhere, the big data paradigm has an insatiable appetite for data – the more the better – and this means feeding the machine with student information, at massive scale, to sustain its operation (Johnson, 2014). They also include eye tracking, automated online dialog analysis, survey data from school ecosystems, log data analysis at individual and collaborative levels, and visual learning analytics applied to IoT data (Nistor and García, 2018).

VLE providers

> [a]llow and track activities, locally, using the Internet of things (e.g., smart phones, smart sensors and other cyber physical devices), and globally, via the internet. . . A wide range of learner behaviors (many implicit or non-obvious, such as those collected via metadata emitted by smart phones) generate rich and vast data-streams, which may be stored on servers controlled or not controlled by the online learning platform. (edX, 2018b)

VLEs maximize data collection, conduct constant assessment, and implement human experiments during the learning process. Besides recording the interactions between learners, professors, peers, and the platform itself, VLEs capture real-time data at an 'unprecedented scale' including every mouse click (Prinsloo & Slade, 2013: 240). As one researcher notes:

> As opposed to the physical classroom, there is a virtual record of every transaction that takes place in the classroom. Every discussion post, every essay, every teacher comment is captured for posterity. . . Big Brother can now see everything. (McCluskey and Winter, 2014)

VLEs use cookies to collect both personal and nonpersonal information, and note that their services may not function if users disable cookie collection (Cousera, 2018; edX, 2018b; Udacity, 2018b). Learners have no meaningful ability to opt-out of these practices. VLE privacy policies note that learners can disable the cookies tracking user behavior, but also state that doing so will limit platforms' functionality – providing no real alternative to complete participation. The primary mechanism VLEs use to cultivate transparency and trust is providing users with an email address that they can send questions to about their information policies or privacy concerns. Learners have minimal ability to access and oversee the accuracy of the assessments, inferences, and resulting credentials created by VLEs that will increasingly affect their academic achievement and employment opportunities. Under American law, they have no say in VLE information practices, aside from the ability to request VLEs delete the information maintained about them.

VLE privacy policies are imprecise and confusing. This uncertainty is compounded by the provision in most VLE privacy policies that permit them to amend material aspects of their data use and protection without requiring user consent and perhaps without user notice. They take continued use of services as a binding acceptance of any changes (Coursera, 2018; edX, 2018b; Udacity, 2018b). The possibility that VLEs might engage in unexpected privacy practices at any time is significant in and of itself, especially when a similar lack of transparency doomed the 100-million Gates Foundation-funded nonprofit inBloom.

Embedded assessment and competency-based credentials. Online platforms involve digital intermediation, which makes every user interaction legible and durable. The ever-observant eye of digital 'instructors' not only captures, but constantly evaluates learner performance during the learning process. MOOC/VLE platforms generate, create, and record continuous embedded assessment of students' progress and proficiency. Computerized systems diagnose student progress, evaluate instructional options, and deliver content based on learners' specific needs. Granular data allows VLE platforms to create epistemologies, learner profiles, and credentials on a concept-by-concept level. Ed tech systems use complex 'learning analytics' to interpret these data to reflect student progress in finer detail than classroom teachers. They keep track of learners' progress in real time, 'embedd[ing] assessment' within instruction (Cope and Kalantzis, 2016; Dede, 2011). VLEs use these up-to-date 'knowledge maps' with predictive algorithmic models to mimic how teachers and guidance counselors 'personalize' instruction and advice to match individual students' progress and needs (Lorch, 2013; Metz, 2013). These data can also help identify which users need help and fuel research about what methods work best and for whom.

Learner profiles are necessarily associated with specific learners, even if platforms use random numbers as identifiers instead of real names (Prinsloo and Slade, 2015). This means that they would fall under the definition of personally identifiable student information protected by FERPA if they had collected information about students from schools, rather than directly from users. Users' education records may also take on undue importance. Many education reformers want to learner profiles for school admission and hiring instead of traditional transcripts, records, and resumes (Craig, 2015). The

Department of Education strongly supports this 'competency-based credentialing' (Office of Postsecondary Education, U.S. Department of Education, 2014).

Research and revenue purposes. VLEs also present their data collection and analysis as a boon to learning science and future generations of students. Practitioners are eager for more data because:

> For the first time, Big Data gives the modern university the tools to separate what is essential from what is accidental. We can now see where we are successful and where we need to improve. We can see where students are having issues and where they are able to grasp and use concepts. (McCluskey and Winter, 2013)

MOOCs hold out the potential of using data collected through their digital interfaces, in a productive feedback loop, to improve their courses – and education overall. They articulate grand hopes that analysis of this previously unimaginable wealth of educational information would lead to breakthroughs in learning science and 'disrupt' traditional pedagogical practices (Lebron and Shahriar, 2015). As noted by Coursera founder Daphne Koller, 'Every action, no matter how inconsequential it may seem, becomes grist for the statistical mill' (Carr, 2012).

This experimentation and observation are essential components of many VLE missions and business models. edX, for example, highlights that its mission is not only to 'present the best of higher education online, offering opportunity to anyone who wants to achieve, thrive, and grow', but also 'to research that will allow us to understand how students learn, how technology can transform learning, and the ways teachers teach on campus and beyond'. Platform researchers, for example, found that users were not completing 10-minute lecture clips, and shifted to using short, interactive videos instead.

VLE privacy policies also give providers the ability to put information that students offer during the learning process to secondary purposes. As a consequence, VLE privacy policies frequently grant providers broad licenses to materials created by enrollees that would allow them to publish student-created content on their sites or place them in marketing materials (Prinsloo and Slade, 2015). Udacity's terms provide that users grant it 'an irrevocable, worldwide, perpetual, royalty-free and non-exclusive license to use, distribute, reproduce, modify, adapt, publicly perform and publicly display such User Content . . . for any purpose (including for any commercial purpose)'(Udacity, 2018c).

Coursera indicates, 'it may publish [personal] information via extensions of [its] Platform that use third-party services, like mobile applications and reserves the right to use non-personalized data for unspecific 'business purposes'. edX asserts ownership of student data in perpetuity. Both current and future edX members have the license to access use of student posts – which may contain student usernames – for their own purposes (Prinsloo and Slade, 2015). Learner data may also be available to the general public, either as VLE marketing material, published research, or open alternative credentialing systems. MOOC/VLE learner data might also be monetized like any other asset in the case of an acquisition or bankruptcy (Lederman, 2014).

VLE information norms undermine prominent education paradigms

Continuing the CI analysis, we examine VLE privacy norms given prominent paradigms of education. This is complicated by the fact that education serves a plurality of purposes in society – both individuals and communities. Accordingly, we incorporate Zeide's (2018) analysis of learning technologies and education paradigms to examine VLE information practices in terms of prioritizing these contextual values, rather than positing a definitive paradigm of education. Drawing on prominent themes in history, jurisprudence, and public discourse, this work considers education as promoting economic, democratic, equitable, and self-actualization purposes. These categories are not meant to be canonical, but to demonstrate the varied and countervailing values and purposes that can shape, and be shaped by, information flow.

Pluralistic education paradigms

Public discourse indicates that people have different views about the essential function of education in society. The economic paradigm considers education as the means to promote national financial prosperity and competitiveness by developing productive workers, harmonizing labor supply and demand, and optimizing education efficiency. Historically, it draws upon the scientific management principles embraced by educations in response to the Industrial Revolution. Economic education purposes emphasize narrow curricula that impart only the necessary information and intervening or guiding students to paths most likely to reach success and achievement. Prominent proponents of ed tech 'disruption' describe this as a goal to 'Hone the skills, capabilities, and attitudes that will help our economy remain prosperous and economically competitive' (Christensen et al., 2010: 1).

The democratic paradigm views education as a tool to ensure the success of America's political enterprise by developing a citizenry capable of self-governance and encouraging robust debatee. Christensen et al. (2010) reference citizen development, saying education should '[f]acilitate a vibrant, participative democracy in which we have an informed electorate that is capable of not being "spun" by self-interested leaders'. They also articulate a rationale that corresponds to the idea of education encouraging robust debate, saying that education should 'Nurture the understanding that people can see things differently – and that those differences merit respect rather than persecution' (Christensen et al., 2010: 1).

The equality paradigm promotes education as a core driver of socioeconomic mobility, fair academic and career outcomes, and a meritocratic opportunity system. The US Department of Education consistently emphasizes the importance of 'ensur[ing] equitable educational opportunities' ('Equity of opportunity', n.d.). Equality is also prominent in educational policies and goals designed to 'close the achievement gap' and the 'digital divide' (Christensen et al., 2010: 1).

The self-actualization paradigm focuses less on the instrumentality of education than on the process itself. It promotes self-actualization and fulfillment through personal growth, intellectual development, and creative exploration. Historically, this paradigm

aligns with the philosophy espoused by John Dewey, who said, 'The aim of education is to enable individuals to continue their education . . . the object and reward of learning is continued capacity for growth' (Christensen et al., 2010: 1). To permit such growth, the model seeks to provide with students the resources to accommodate diverse intellectual exploration, the flexibility to pursue individual and 'impractical' interests, and the suspension of judgment and forgiveness of mistakes necessary to encourage experimentation.

The disruptors talk about education as a way to 'Maximize human potential' (Christensen et al., 2010: 1). This is no doubt true, but sufficiently vague that it glosses over the tensions between maximizing human potential in terms of employment rates, future salary, or creative expression. 'Maximizing potential' as a broad purpose does not address the tension between socioeconomic mobility and streamlining efficiency and economic progress by promoting and hiring learners who are already best prepared and predicted most likely to succeed.

VLE ideals and less lofty realities

VLEs originally articulated ideals that corresponded with the values of these various education paradigms (McGregor, 2013). Udacity's original mission furthered multiple education paradigms: to bring 'accessible, affordable, engaging, and highly effective higher education to the world', calling higher education 'a basic human right' (About Us [Udacity]).

Today, however, VLEs focus on providing career certifications for a fee, rather than liberal arts academic courses (Bernhard and Klein, 2015). Their business models, pedagogical approaches, and accordingly, their information flow aligns with education as a means of individual achievement and career advancement (Young, 2016). VLEs still present their platforms as a means to self-actualization, but a concept of self-actualization narrowed with credential acquisitions and career advancement (Longstaff, 2017: 315):

> In return for your hard work, Udacity offers a range of certification options that are recognized by major technology companies who are actively recruiting from the Udacity student body. Join the hundreds of thousands of Udacity students who have already been empowered by this new form of learning. (Udacity, 2018a)

The MOOC/VLE worldview promotes education as a means to achieve a very narrow set of outcomes: primarily as a means to be more competitive in the labor market. Despite rhetoric extolling education as a means to promote democracy, equality, and self-actualization, most online education providers offer educational experiences focused on the economic value of acquiring marketable skills and credentials. Their systems prioritize easily quantifiable competencies, rather than intangible, 'soft' skills like critical thinking and self-regulation. They 'streamline' curricula, automate instruction, and experiment on learners for the sake of efficiency and optimization. Virtual learning platforms present education as merely a series of learning outcomes to be met.

Streamlined academic and career attainment. Today, VLEs focus on education's instrumental value. 'Empowerment', for Udacity, is no longer tied to human rights. Instead, the platform wants 'students to not just advance their education, but to land their dream job in technology through a relevant 21st century education' (Shieber, 2014). VLEs advertise their services as a more cost- and time-efficient means to obtain skills and credentials valuable in the labor market, employer anyone to advance their education and career (Udacity, 2018a). Efficacy, according to VLEs, is about achieving career goals – Udacity even offers a money-back guarantee if learners do not land a job within 6 months of completing coursework.

Both democratic and egalitarian ideals emphasize the importance of ameliorating inequality by providing broad access to affordable, quality education. Since *Brown v. Board of Ed.* in 1954, education has also evolved as a primary means by which we seek to ensure the equality inherent in a democratic society (*Brown v. Board of Education*, 1954). As the Supreme Court noted in *Grutter v. Bollinger* (2003), 'the diffusion of knowledge and opportunity through public institutions of higher education must be accessible to all individuals regardless of race or ethnicity'. However, access alone may not increase equitable outcomes in practice. Charging fees for certification may also preclude those most in need of credentials from obtaining them. Research suggests that VLEs are most useful for already successful students, who consistently fare better than their less accomplished classmates (Reich and Ho, 2014). VLEs, for example, notoriously failed – as did the students – at San Jose Community College (Kolowich, 2013b).

An overly narrow focus on credential attainment may preclude civic education and broad exposure to ideas central to the democratic education paradigm. Focusing on the instrumental value of education to obtain labor market credentials may create a rational focus on completing formal requirements rather than engaging in the challenges that accompany deep learning (Labaree, 1997). This is in sharp contrast to the education paradigms that value more abstract and integrated human development, particularly regarding equality and self-actualization. For example, Mireille Hildebrandt (2017) notes that we want education to promote critical thinking not only with respect to particular issues, but for "reflection *on what and on how* we learn".

Data-defined metrics and outcomes. Making education more 'effective' (Udacity, 2018a) is a mixed bag; it depends on what is being optimized. Data derived-metrics and data-defined outcomes may undermine democratic and self-actualization ideals. Absent technical, institutional, or policy protocols, the technology of data-driven systems will have considerable influence on what counts as 'learning', 'academic attainment', and 'education'. In VLEs, technological affordances confine pedagogical decisions to information that can be captured, processed, and preserved by computer.

Many characteristics traditionally considered in physical classrooms are not commensurable. Despite a detailed clickstream of data, the platforms can't tell (yet) when a student is sick or socially isolated. They also can't account for intangibles like creativity, independence, and resilience. This undermines the goals of the democratic which value the critical thinking skills and self-actualization paradigms which embrace intellectual freedom. VLE business incentives may also create perverse incentives to devote resources to the learners most likely to succeed in order to boost success rates (Patterson, 2016;

Zeide, 2017a). Strict focus on quantifiable goals and efficiency under the economic model of education may lead to better statistical outcomes at the expense of individual attainment, and disparately impact learners on the wrong side of the learning gap. Recently, for example, a university president sought to use student data to determine which students were most likely to drop out, and then encourage them to do so early in the semester to improve the schools' retention statistics (Jaschik, 2016).

Improving student outcomes through data mining may not, in fact, promote better pedagogy, but simply streamline the system regarding measured outcomes. Adaptive technology and predictive models would provide a 'rational' and 'scientific' basis to reduce the risk and costs of student failure and streamline stratification. There is growing concern that big data techniques may unintentionally have a discriminatory impact. Instead of 'teaching to the test', these systems will be adaptive to optimize the metrics used as proxies for competencies. As Campbell's Law sagely notes, 'The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor'.

Digital mediation and data maximization. VLEs' ubiquitous surveillance and maximized data collection undermine many values in the noneconomic paradigm. Learners worried about the consequences of their actions in classrooms suffer chilling, conforming, and credentialing effects. Chilling effects occur when surveillance reduces learners' participation and expression in both classroom activities and the political activism often found in higher education environments. Potential surveillance effects are particularly pernicious under the democratic paradigm for the potential to discourage civic participations, free thought, and robust debate. As noted above, Neil Richard's (2008) scholarship highlights the importance of 'intellectual privacy' in promoting robust debate. The Supreme Court's jurisprudence also highlights the necessity of privacy to ensure intellectual and social freedom. As Margot Kaminski (2017) notes, prominent theories supporting First Amendment protection 'also justify the protection of a broader array of intellectual and social freedoms'. It may have a conforming effect and prompt learners to move their viewpoints closer to that of the perceived mainstream – unconsciously (Kaminski and Witnov, 2015). Who will play devil's advocate for an unpopular cause if doing so will go down in a possibly permanent record? This runs counter to the prominent value of education in cultivating a citizenry capable of self-governance (Westheimer and Kahne, 2004).

Constant surveillance and assessment may also undermine self-actualization aims. As such surveillance becomes more ubiquitous with 'anywhere, anytime' learning, it may also constrain the unselfconscious play that facilitates the curiosity, intellectual exploration, and expression consistent with human flourishing. Julie Cohen (1995), for example, has noted the potential for digital reading platforms to undermine user intellectual autonomy and alter their intellectual exploration.

Continuous evaluation and automated personalization. While automated instruction may be easier to scale, the implications of data-driven 'personalization' are murky at best. Reformers and providers promise big data-driven education as a low-cost way to improve individual student performance and engagement. These tools may help close the

'achievement gap' by providing more individual attention to struggling and underserved students. However, the existing evidence about the success of personalized learning platforms is inconclusive.

As the Supreme Court noted in *Keyishian v. Bd. of Regents*, the democratic model of education prioritizes exposing students to a 'robust exchange of ideas'. Personalized instruction may also create a filter bubble that undermines this goal. It may also undermine self-actualization aims. There will be less exposure to new ideas, and less room for serendipitous discovery of novel interests. In terms of self-actualization, personalized learning tools allow more individualization of learning paths and content. In this way, they support students' unique abilities, cognitive approaches, and personal interests. To the degree this permits students to carve their own intellectual paths and pursue personal interests, this supports the discovery model of education. While personalized learning systems can cater more to learners' preferences more than one-size-fits-all lectures, they do not offer unbounded intellectual exploration (Watters, 2011).

Using learner profiles as tests and transcripts may be more 'efficient' than the traditional separation between formative and summative assessment. This reflects the norm separating formative assessment – used as diagnostic tools to inform the student and teaching – and summative assessments used to evaluate student performance. As a result, data-driven Ed tech may increase students' sense of vulnerability, in ways that will have a detrimental effect on their ultimate performance and equitable outcomes. Although professors traditionally collected 'data' about students throughout a course, they typically suspended judgment by delaying summative assessment until the end of a semester. The separation of summative and formative assessment gave slow-starters a chance to catch up with their peers, promoting meritocracy and equality. This allows students to develop and demonstrate ability over time, permitting slow starters the space to catch up to others.

In VLEs, learner profiling and embedded assessment collapses formative feedback, summative assessment, and credentialing creation – every action becomes part of learners' algorithmic transcripts. A sense of heightened consequence may have what Zeide (2016b) calls a 'credentialing effect' that demotivates and ultimately diminishes the performance of struggling students. Research suggests that students' sense of vulnerability impedes academic promise and disproportionately affects minorities (Prinsloo and Slade, 2015).

Student privacy norms also reduce the potential for inaccurate, biased, or outdated records to unfairly limit students' future opportunities. Education privacy norms implicitly recognize that students are not static objects, but evolving individuals shaped as much by their experiences in learning environments as the specifics of instruction an assessment. In creating a barrier to promiscuous disclosure, they allow students to move beyond past mistakes.

Universities have been regarded as 'zones where faculty and students can feel free to pursue any idea, any line of argument, and any intellectual pursuit that they wish- free of the constraints of political pressure, cultural convention, or material interests' (Labaree, 2017: 188). There is also a broader sense of isolation from long-term consequences reflecting higher education institutions as ivory towers. Ignoring the potential pejorative implications of the term, it connotes space for exploration and experimentation without

fear of future consequences. Under student privacy norms, long-term collection of information and its repurposing beyond the educational framework limits the degree of reinvention that may be possible for students.

Data-driven ed tech, in contrast, is currently configured to crate the algorithmic equivalent of the proverbial permanent record. This runs counter to the consistent theme in American society, economic policy, and political rhetoric that the past should not unduly limit future opportunities. Jack Balkin has argued US constitutional history itself is crucially founded on a forward-looking redemptive commitment. Like the expungement of juvenile criminal records or old bankruptcy proceedings, the practical obscurity of classroom proceedings promotes the value Andrew Tutt (2014) calls "revisability". This, in turn, supports individual autonomy.

Private actors and priorities

VLEs' *caveat emptor* privacy norms lie in sharp contrast to the relative safety created by information practices in traditional education environments – and it is likely most VLE users think educational privacy norms apply. VLE websites and marketing materials obscure the fact that they are (predominantly) for-profit entities ultimately responsible to provide return on venture capitalist or shareholder investments – not decide based on what's best for students. Coursera even uses a '.org' domain rather than the '.com' traditional for commercial entities. A lack of transparency and overall uncertainty about VLE information practices also undermines the intellectual safety of these learning environments.

The disclosure, or potential disclosure, of student information to institutional and parties outside the learning environment for noneducational purposes undermines traditional norms of confidentiality in classroom settings. VLE privacy policies permit them to share learner information broadly – possibly business partners, advertisers, purchasers, researchers, and the general public. Learner information that might previously have been irrelevant to outside and non-educational interests now has enormous potential value. The presumption that information generated during the learning process will benefit the data subject's education is also no longer the case.

In the past 50 years, for-profit educators have consistently been called out for prioritizing revenue generation over students' academic interests. A recent congressional report revealed that many profit-seeking higher education institutions allocated most of their resources to marketing and salesman pushing predatory student loans. The technology sector pressures software and platform providers to prioritize short-term profits and growth over other considerations. Startups need to generate revenue – or their potential to do so – to receive additional rounds of funding or pique the interest of deep-pocketed acquirers. Absent self-restraint, VLEs could conceivably sell or share information from students' cognitive maps to insurance companies or future employers without learner knowledge or consent, as Coursera suggested in an early business model (*The Chronicle of Higher Education*, 2012: 40–41).

The boom-and-bust, iterate-and-failure nature of start-up culture – what Zeide (2017b) has called 'beta education' – is problematic in education. Learning is not only a high stakes enterprise for individuals and society, but one which depends on user trust and

intellectual safety for success. New learning providers may not, however, prioritize students' long-term interests, or even commit to operating as educational institutions, as demonstrated by Alt-School's sudden shuttering of two physical primary schools following an investor-oriented 'pivot' (Herold, 2017).

Education versus commercial marketplace: Straddling irreconcilable contexts

As discussed in the previous section, VLEs present an internal contradiction. On one hand, they aspire to the ranks of educational institutions, in the functions they perform, the values they espouse, and the goals they seek to accomplish. Coursera founder, Daphne Koller, has stated that the company's decisions are driven by 'what's best [for] the students' – notwithstanding its for-profit status. (Myers, 2012) Udacity's website similarly states '[W]e put you, the student, at the center of the universe'. (Udacity, 2018a) As education researcher Emily Longstaff (2017) observes, 'Coursera, edX, and Udacity have positioned themselves as a social good', with collective aims 'to promote anywhere, anytime learning; to increase access to world-class education; and connect and empower learners'. However, as described above, VLEs explicitly eschew the heightened protections provided in publicly funded learning environments. VLEs purport to offer educational services for educational purposes, but treat their learners as consumers rather than students, offering the minimal and procedural protection of the commercial notice-and-choice regulatory regime. VLE data practices may undermine their explicit mission through the chilling, conforming, and credentialing effects of constant surveillance, data maximization, embedded assessment, and record retention. Despite rhetoric articulating their commitment to democratic and individual flourishing, they focus on a narrow economic-oriented paradigm. In doing so, they reduce the likelihood that they will achieve the broad aims that are characteristic of the educational context. It is troubling, therefore, that as VLEs are increasingly important in providing for society's educational needs, their data practices undermine the values and ends of an educational context, which are, simultaneously to promote economic and democratic ends while sustaining values of equality and self-actualization.

John Dewey said,

> the social role of education in a democratic society is at once to insure equal liberty and equal opportunity to differing individuals and groups, and to enable the citizens to understand, appraise, and redirect forces, men, and events as these tend to strengthen or to weaken their liberties. (Dewey, 2004).

VLEs, however, see their role as commercial, rather than social, and their services a product, rather than an institution – and their privacy norms reflect and promote these paradigms. As Zeide (2017b) has noted elsewhere 'big data-driven tools define what "counts" as education by mapping the concepts, creating the content, determining the metrics, and setting desired learning outcomes of instruction' (p. 164). If VLEs seek to fill the role of schools, they should adopt privacy practices and information norms that

reflect the magnitude of that public responsibility, and truly foster the plurality of educational values they espouse.

Acknowledgements

We are grateful to organizers of the Conference on Big Data and Education: Ethical and Policy Challenges, Haifa Center for Law and Technology and Center for Ethics and Education, for inviting us to present this work, for although we have together and separately presented it elsewhere, this provided a final push to bring a years' long project to fruition. Audiences at other venues, too, have greatly enhanced this work; they include NYU's Privacy Research Group, Annenberg School for Communication, University of Pennsylvania, Symposium on Digital Ethics, Loyola University, and Symposium on Building Privacy into Data-Driven Education, Microsoft Policy & Innovation Center.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We gratefully acknowledge support from US-AFSOR: ONR BAA 07-03 (MURI), US NSF CNS-0831124 and SES-1537324, Intel Science and Technology Center for Social Computing, and Microsoft Corporation.

References

Anders G (2013) Sebastian Thrun's online goal: Act where college isn't working. *Forbes*, 3 April. Available at: https://www.forbes.com/sites/georgeanders/2013/04/03/sebastian-thruns-online-goal-act-where-college-isnt-working/

Banks C and Meinert E (2016) *The Acceptability of MOOC Certificates in the Workplace*. International Association for Development of the Information Society. Available at: https://files.eric.ed.gov/fulltext/ED571496.pdf

Bernhard MP and Klein MA (2015) Provost Lauds HarvardX, but questions its financial sustainability. *The Harvard Crimson*, 26 October. Available at: http://www.thecrimson.com/article/2015/10/26/garber-edX-financial-sustainability/

Bolkan J (2015) EdX, Arizona state U add online proctoring service for global freshman academy. *Campus Technology*, May 18. Available at: http://campustechnology.com/articles/2015/05/18/edx-arizona-state-u-add-online-proctoring-service-for-global-freshman-academy.aspx

Briones J (2018) The data quality campaign's legislative summary and analysis. *Data Quality Campaign*. Available at: https://2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2018/09/2018-DQCLegislative-Summary.pdf

Brown v. Board of Education, 347 US 483 (1954).

Byrnes N (2015) Uber for Education. *MIT Technology Review*. Available at: http://www.technologyreview.com/news/539106/uber-for-education/ (accessed 27 July 2018).

Carey K (2015) *The End of College: Creating the Future of Learning and the University of Everywhere*. New York: Riverhead Books.

Carr N (2012) The crisis in higher education. *MIT Technology Review* 115(6): 32–40.

CB Insights (2016) Mega-rounds boost global Ed tech funding to new record. *CB Insights*, 19 January. Available at: https://www.cbinsights.com/blog/2015-global-ed-tech-funding/

Christensen C, Johnson CW and Horn MB (2010) *Disrupting Class, Expanded Edition: How Disruptive Innovation Will Change the Way the World Learns*, 5th edn. New York: McGraw-Hill.

Cohen JE (1995) A right to read anonymously: A closer look at copyright management in cyberspace. *Connecticut Law Review* 28: 981.

Cope B and Kalantzis M (2016) Big data comes to school: Implications for learning, assessment, and research. *AERA Open* 2: 1–19.

Consortium for School Networking (CoSN) (2014) *Protecting Student Privacy in Connected Learning Toolkit*. Washington, DC: CoSN. Available at: http://www.cosn.org/focus-areas/leadership-vision/protecting-student-privacy

Coursera (2018). Privacy policy. Available at: https://www.coursera.org/about/privacy

Craig R (2015) Coursera, Udacity and the future of credentials. *Forbes*. Available at: http://www.forbes.com/sites/ryancraig/2015/09/30/coursera-udacity-and-the-future-of-credentials/ (accessed 30 September 2018).

Craig R (2016) Competency-based hiring: 10 signals the shift is happening. *Forbes*. Available at: http://www.forbes.com/sites/ryancraig/2016/10/01/competency-based-hiring-10-signals-the-shift-is-happening/ (accessed 1 October 2018).

Cuban L (2013) *Inside the Black Box of Classroom Practice: Change without Reform in American Education*. Cambridge, MA: Harvard Education Press.

Dede C (2011) Emerging technologies, ubiquitous learning, and educational transformation. In: Kloos CD, Gillet D, García RMC, et al. (eds) *Towards Ubiquitous Learning*. Berlin; Heidelberg: Springer. pp. 2–3, 7.

'De-identified data' (n.d.) U.S. Department of Education. Available at: https://studentprivacy.ed.gov/sites/default/files/resource_document/file/data_deidentification_terms.pdf

Dewey J (2004) *Democracy and Education*. Chelmsford, MA: Courier Corporation.

edX (2018a) About us. Available at: https://www.edx.org/about-us (accessed 26 May 2015).

edX (2018b) Privacy policy. Available at: https://www.edx.org/privacypolicy (accessed 19 August 2013).

'Equity of opportunity' (n.d.). U.S. Department of Education. Available at: https://www.ed.gov/equity

Family Educational Rights and Privacy Act (FERPA), 20 U.S. Code § sec. 1232g (1974).

Fidanza M (2015) Aging out of in loco parentis: Towards reclaiming constitutional rights for adult students in public schools. *Rutgers Law Review* 67: 805.

Fitzgerald B (2014) The day and the data; catching policy up to reality. *FunnyMonkey*. Available at: http://funnymonkey.com/blog/day-and-data-catching-policy-reality (accessed 30 January 2018).

Ghilay Y (2017) *Online Learning in Higher Education*. Hauppauge, NY: Nova Science Publishers.

Grutter v. Bollinger, 539 US 306 (2003).

Herold B (2014) 'Big data' research effort faces student-privacy questions. *Education Week*, 22 October. Available at: https://www.edweek.org/ew/articles/2014/10/22/09learnsphere.h34.html

Herold B (2017) Citing investor priorities, altschool closes some sites, pivots to software sales. *Education Week*. https://blogs.edweek.org/edweek/DigitalEducation/2017/11/AltSchool_investors_closes_sites_software_Sales.html

Hildebrandt M (2017) Learning as a machine: Crossovers between humans and machines. *Journal of Learning Analytics* 4(1): 6–23.

Ho AD, Reich J, Nesterko SO, et al. (2014) HarvardX and MITx: The first year of open online courses, Fall 2012-summer 2013. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2381263

Hollands FM and Tirthali D (2014) Why do institutions offer MOOCs? *Online Learning: Official Journal of the Online Learning Consortium* 18(3): 6–13.

Hutt E (2016) A brief history of the student record: A paper for the Asilomar conference on student data and records in the digital era. *Ithaka S+R*. Available at: http://sr.ithaka.org/?p=283886

Jaschik S (2016) Furor at Mount St. Mary's over president's alleged plan to cull students. *Inside Higher Ed*. Available at: https://www.insidehighered.com/news/2016/01/20/furor-mount-st-marys-over-presidents-alleged-plan-cull-students (accessed 20 January 2018).

Johnson JA (2014) The ethics of big data in higher education. *International Review of Information Ethics* 7(1): 3–10.

Kalman YM (2014) A race to the bottom: MOOCs and higher education business models. *Open Learning: The Journal of Open, Distance and E-learning* 29(1): 5–14.

Kaminski M (2017) Intellectual and social freedom. In: Gray D and Henderson S (eds) *The Cambridge Handbook of Surveillance*. Cambridge: Cambridge University Press. pp. 465–468.

Kaminski ME and Witnov S (2015) The conforming effect: First amendment implications of surveillance, beyond chilling speech. *University of Richmond Law Review* 49: 55.

Keyishian v. Board of Regents of Univ. of State of NY, 385 US 589 (1967).

Knewton–Education datapalooza (2012) Washington, DC. Available at: http://www.youtube.com/watch?v=Lr7Z7ysDluQ&feature=youtube_gdata_player (accessed 4 November 2018).

Koller D (2013) MOOCs can be a significant factor in opening doors to opportunity. *EdSurge*. Available at: https://www.edsurge.com/news/2013-12-31-daphne-koller-moocs-can-be-a-significant-factor-in-opening-doors-to-opportunity (accessed 31 December 2017).

Kolowich S (2012) Elsevier partners with edX to provide free versions of textbooks to MOOC students. *Inside Higher Ed*. Available at: https://www.insidehighered.com/news/2012/09/26/elsevier-partners-edx-provide-free-versions-textbooks-mooc-students

Kolowich S (2013a) In deals with 10 public universities, Coursera bids for role in credit courses. *The Chronicle of Higher Education*. Available at: https://www.chronicle.com/article/In-Deals-With-10-Public/139533 (accessed 30 May 2018).

Kolowich S (2013b) San Jose state U. Puts MOOC project with Udacity on hold. *The Chronicle of Higher Education*. Available at: http://chronicle.com/article/San-Jose-State-U-Puts-MOOC/140459 (accessed 19 July 2018).

Labaree DF (1997) Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal* 34(1): 39–81.

Labaree DF (2017) *A Perfect Mess: The Unlikely Ascendancy of American Higher Education*. Chicago, IL: The University of Chicago Press.

Leber J (2012) The technology of massive open online courses. *MIT Technology Review*. Available at: https://www.technologyreview.com/s/506326/the-technology-of-massive-open-online-courses/ (accessed 5 November 2017).

Lebron D and Shahriar H (2015) Comparing MOOC-based platforms: Reflection on pedagogical support, framework and learning analytics. In: *2015 International Conference on Collaboration Technologies and Systems (CTS)*. Available at: https://ieeexplore.ieee.org/document/7210417

Leckart S (2012) The Stanford education experiment could change higher learning forever. *Wired*. Available at: https://www.wired.com/2012/03/ff_aiclass/

Lederman D (2014) FTC warns of student privacy risk in ConnectEDU bankruptcy. *Inside Higher Ed*. Available at: https://www.insidehighered.com/quicktakes/2014/05/27/ftc-warns-student-privacy-risk-connectedu-bankruptcy

Longstaff E (2017) How MOOCs can empower learners: A comparison of provider goals and user experiences. *Journal of Further and Higher Education* 41(3): 314–327.

Lorch K (2013) EdX to examine personalized MOOC experience. *The Harvard Crimson*. Available at: https://www.thecrimson.com/article/2013/10/17/harvardx-to-examine-personalized-mooc-16 Theory and Research in Education experience/

Lu A (2013) Student data privacy addressed by states. *The Pew Charitable Trusts*. Available at: http://www.pewstates.org/projects/stateline/headlines/protecting-student-privacy-in-the-data-age-85899527146 (accessed 17 December 2017).

MacCarthy M (2014) Student privacy: Harm and context. *International Review of Information Ethics* 21: 11–24.

McCluskey F and Winter M (2013) *The Idea of the Digital University: Ancient Traditions, Disruptive Technologies and the Battle for the Soul of Higher Education*, vol. 16.

McCluskey F and Winter M (2014) Academic freedom in the digital age. *On the Horizon* 22(2): 136–146.

McGregor K (2013) Will MOOCs help to democratise higher education? *University World News*. Available at: http://www.universityworldnews.com/article.php?story=20130622164019140

Madden M, Cortesi S, Gasser U, et al. (2012) Parents, teens, and online privacy. *Pew Internet & American Life Project* 9: 8.

Marshall S (2014) Exploring the ethical implications of MOOCs. *Distance Education* 35(2): 250–262.

Metz R (2013) Sebastian Thrun on the Future of Learning. *MIT Technology Review*. Available at: https://www.technologyreview.com/s/517181/sebastian-thrun-on-the-future-of-learning/

Murray J (2017) Has the time come for personalized higher education? *Brookings*. Available at: https://www.brookings.edu/blog/brown-center-chalkboard/2017/02/21/has-the-time-come-for-personalized-higher-education/

Myers CB (2012) Coursera raises $16 million to bring free online education to millions. *The Next Web*. Available at: https://thenextweb.com/insider/2012/04/18/coursera-raises-16-million-and-partners-with-top-tier-universities-to-bring-free-online-education-to-millions/

New J (2013) Partnership gives students access to a high-price text on a MOOC budget. *The Chronicle of Higher Education*. Available at: http://chronicle.com/article/Partnership-Gives-Students/139109/

Nissenbaum H (2009) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford Law Books.

Nistor N and García ÁH (2018) What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior* 89: 335–338.

Office of Postsecondary Education, U.S. Department of Education (2014, December 18) Competency-Based Education Programs–Questions and answers. Available at: https://ifap.ed.gov/dpcletters/attachments/GEN1423.pdf

O'Reilly U-M and Veeramachaneni K (2014) Technology for mining the big data of MOOCs. *Research & Practice in Assessment* 9(2): 29–37.

Pappano L (2012) The year of the MOOC. *The New York Times*. Available at: http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

Patterson M (2016) Colleges v. corporations: The sad tale of Mount St. Mary's. *America*. Available at: http://www.americamagazine.org/content/all-things/colleges-v-corporations-sad-tale-mount-st-marys

PBS Newshour (2013) How free online courses are changing the traditional liberal arts education. *PBS Newshour, 8* January. Available at: https://www.pbs.org/newshour/show/how-free-online-courses-are-changing-traditionaleducation

Perez-Kudzma C (2007) Fiduciary duties in academia: An uphill battle. *Idea* 48: 491.

Pesare E, Roselli T, Rossano V, et al. (2015) Digitally enhanced assessment in virtual learning environments. *Journal of Visual Languages and Computing* 31: 252–259.

Prinsloo P and Slade S (2013) An evaluation of policy frameworks for addressing ethical considerations in learning analytics. In: *Proceedings of the third international conference on learning analytics and knowledge*, pp. 240–244. New York: ACM. Available at: https://dl.acm.org/citation.cfm?id=2460344

Prinsloo P and Slade S (2015) Student privacy self-management: Implications for learning analytics. In: *Proceedings of the fifth international conference on learning analytics and knowledge*, pp. 83–92. ACM. Available at: http://dl.acm.org/citation.cfm?id=2723585

Protection of Pupil Rights Amendment, 20 § 1232h (1978).

Reich J and Ho A (2014) The tricky task of figuring out what makes a MOOC successful. *The Atlantic*. Available at: http://www.theatlantic.com/education/archive/2014/01/the-tricky-task-of-figuring-out-what-makes-a-mooc-successful/283274/

Richards NM (2008) Intellectual privacy. *Texas Law Review* 87: 387.

Selwyn N (2014) *Distrusting Educational Technology: Critical Questions for Changing Times*. New York; London: Routledge; Taylor & Francis.

Shah D (2016a) MOOC trends in 2016: College credit, credentials, and degrees. Available at: https://www.class-central.com/report/mooc-trends-credit-credentials-degrees/

Shah D (2016b) Monetization over massiveness: Breaking down MOOCs by the numbers in 2016. *EdSurge*. Available at: https://www.edsurge.com/news/2016–12–29-monetization-over-massiveness-breaking-down-moocs-by-the-numbers-in-2016

Shelton v. Tucker, 364 US 479 (1960).

Shieber J (2014) Its audacity undiminished, Udacity raises $35 million to train a new generation of developers. *TechCrunch*. Available at: http://techcrunch.com/2014/09/24/its-audacity-undiminished-udacity-raises-35-million-to-train-a-new-generation-of-developers/

Singer N (2015) Online test-takers feel anti-cheating software's uneasy glare. *The New York Times*. Available at: http://www.nytimes.com/2015/04/06/technology/online-test-takers-feel-anti-cheating-softwares-uneasy-glare.html

Student Online Personal Information Protection Act (SOPIPA) (2014), CAL. BUS. & PROF. CODE §§ 22584–22585.

Student Privacy Pledge (2016) *Future of Privacy Form and Software Pledge to Parents & Students*. Washington, DC: Student Privacy Pledge. Available at: http://studentprivacypledge.org/

Sweezy v. New Hampshire, 354 US 234 (1957).

The Chronicle of Higher Education (2012) Coursera Michigan contract. *The Chronicle of Higher Education*. Available at: http://chronicle.com/article/Document-Examine-the-U-of/133063/ (accessed 17 April 2018).

The Economist (2017) The return of the MOOC: Established education providers v new contenders. *The Economist*, 12 January. Available at: https://www.economist.com/special-report/2017/01/12/established-education-providersv-new-contenders

Tutt A (2014) The revisability principle. *Hastings Law Journal* 66: 1113.

Udacity (2018a) About us. Available at: https://www.udacity.com/us (accessed 28 June 2015).

Udacity. (2018b) Privacy policy. Available at: https://www.udacity.com/legal/privacy (accessed 1 July 2015).

Udacity (2018c), Terms of Service. Available at: https://www.udacity.com/legal/terms-of-use.

Vance A (2016) Data privacy laws follow lead of Oklahoma and California. *The State Education Standard* 16(2): 25.

Voss BD (2013) Massive Open Online Courses (MOOCs): A primer for university and college board members. Available at: http://agb.org/sites/agb.org/files/report_2013_MOOCs.pdf

Wan T (2016) Ka'Ching! 2016 US Edtech funding totals $1 billion. *EdSurge*. Available at: https://www.edsurge.com/news/2016-12-26-ka-ching-2016-us-edtech-funding-totals-1-billion

Watters A (2011) Pearson and Knewton: Big data and the promise of personalized learning. *Hack Education*. Available at: http://hackeducation.com/2011/11/01/pearson-and-knewton-big-data-and-the-promise-of-personalized-learning

Watters A (2014) Teaching machines: A brief history of 'Teaching at Scale'. *Hack Education*. Available at: http://hackeducation.com/2014/09/10/teaching-machines-teaching-at-scale

Watters A (2015) The first teaching machines–The history of the future of education. *Medium*. Available at: http://hackeducation.com/2015/02/03/the-first-teaching-machines

Westheimer J and Kahne J (2004) What kind of citizen? The politics of educating for democracy. *American Educational Research Journal* 41(2): 237–269.

White B (2007) Student rights: From in loco parentis to sine parentibus and back again? Understanding the Family Educational Rights and Privacy Act in higher education. *Brigham Young University Education and Law Journal*. Available at: https://digitalcommons.law.byu.edu/elj/vol2007/iss2/6/

Wieman v. Updegraff, 344 US 183 (1952).

Young E (2015a) Educational privacy in the online classroom: FERPA, MOOCs, and the big data conundrum. *Harvard Journal of Law & Technology* 28(2): 549–593.

Young JR (2016) Are MOOCs forever? *The Chronicle of Higher Education*. Available at: http://www.chronicle.com/article/Are-MOOCs-Forever-/237130/

Young JR (2013) *Beyond the MOOC Hype: A Guide to Higher Education's High-tech Disruption*. Washington, DC: The Chronicle of Higher Education.

Young JR (2015b) Meet the new, self-appointed MOOC Accreditors: Google and Instagram. *The Chronicle of Higher Education Blogs: Wired Campus*. Available at: http://chronicle.com/blogs/wiredcampus/meet-the-new-self-appointed-mooc-accreditors-google-and-instagram/55807

Zeide E (2016a) Student privacy principles for the age of big data: Moving beyond FERPA and FIPPs. *Drexel Law Review* 8(2): 339.

Zeide E (2016b) The credentialing effect: Psychological effects of ubiquitous capture and constant evaluation. In: *Privacy Law Scholars Conference*, Washington, DC., 2-3 June 2016.

Zeide E (2017a) The limits of education purpose limitations. *University of Miami Law Review* 71(2): 493–527.

Zeide E (2017b) The structural consequences of big data-driven education. *Big Data* 5(2): 164–172.

Zeide E (2018) Learning v. Education: How technology shapes purposes and paradigms. Available at SSRN: https://ssrn.com/abstract=3289733

Author biographies

Elana Zeide is a PULSE Fellow in AI, Law & Policy at UCLA Law School's Program on Understanding Science, Law, and Evidence.

Helen Nissenbaum is a Professor of Information Science at Cornell Tech.